

UN ESTUDIO DE SIMULACIÓN ACERCA DEL ERROR DE TIPO I EN LA DETECCIÓN DEL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM

MARÍA ESTER AGUERRI

Resumen de la Tesis de Magister Scientiae en Biometría defendida el 24 de mayo de 2000

El funcionamiento diferencial del ítem (DIF) se presenta cuando sujetos de un mismo nivel de habilidad pero de diferentes grupos tienen distinta probabilidad de contestar correctamente al ítem. El objetivo de la presente tesis es investigar los factores influyentes en la detección del funcionamiento diferencial del ítem cuando en realidad no lo tiene. Los factores cuyo efecto se investiga son el método utilizado para la detección del DIF, el tamaño de muestra, la distribución de la habilidad y los parámetros del ítem. Se simuló las respuestas de sujetos a una prueba de 20 ítems con el modelo logístico de tres parámetros. Se consideraron tres posibilidades para la distribución de la habilidad y dos tamaños de muestra. Los parámetros de los ítems combinan cuatro niveles del parámetro de discriminación con cinco niveles del parámetro de dificultad. Se utilizaron dos métodos no paramétricos: el χ^2 de Mantel-Haenszel (MH) y el que se deriva del criterio establecido por el Educational Testing Service (ETS) para la clasificación de los ítems por su DIF, y la prueba normal para la diferencia de los parámetros de dificultad (BLM), en el marco de la Teoría de Respuesta al Ítem. Se destaca el método BLM, porque con dicho método la proporción de DIF erróneo se mantuvo dentro del nivel de significación elegido para la prueba en 116 de las 120 combinaciones de los niveles de los factores. Con este método se cometió Error de Tipo I más allá del nivel de significación de la prueba en ítems difíciles y de discriminación que supera al menor nivel, cuando el grupo minoritario pertenece a una población en desventaja en cuanto a la habilidad. Situación que analizada en detalle se puede adjudicar a la insuficiencia de observaciones en los niveles altos de la habilidad por lo que la inexactitud de las estimaciones conduce a detectar un falso DIF. Tanto el método MH como el método ETS presentan una gran variabilidad en cuanto a la proporción de DIF erróneo. Aunque el método ETS se mantiene dentro del nivel de significación de la prueba en 90 de las 120 combinaciones de los niveles de los factores, el método MH sólo lo logra en 22 de tales oportunidades. El método MH alcanza los valores más altos de proporción de DIF erróneo para los ítems de discriminación medio-alta y alta y considerados muy fáciles, o fáciles o muy difíciles cuando uno de los grupos proviene de una población en marcada desventaja en cuanto a la habilidad, y en el caso opuesto, cuando uno de los grupos pertenece a una población con marcada ventaja en cuanto a la habilidad, el ítem que presenta mayor proporción de DIF erróneo fue un ítem de dificultad intermedia y de la más alta discriminación. En cuanto al método ETS, se observa que los valores más altos de proporción de DIF erróneo corresponden a ítems de discriminación medio-alta y alta y considerados muy fáciles cuando uno de los grupos proviene de una población en marcada desventaja en cuanto a la habilidad, y en el caso opuesto, cuando uno de los grupos pertenece a una población con marcada ventaja, el ítem que presenta mayor proporción de DIF erróneo fue un ítem que es fácil y de la más alta discriminación. Dada la naturaleza de los datos no se logró satisfacer los supuestos necesarios para la aplicación de un análisis de la varianza que permitiera concluir sobre la totalidad de los factores involucrados en esta tesis. Se rechazó la independencia del método con la ocurrencia del Error de Tipo I en los tres niveles planteados para la habilidad. La tendencia que se manifiesta es a cometer Error de Tipo I cuando el método utilizado es el MH, y a no cometerlo cuando el método es el BLM, para todos los niveles de la habilidad. Mientras que,

el método ETS muestra una marcada tendencia a no cometer Error de Tipo I cuando el grupo focal es aventajado por el grupo de referencia o están en igualdad de condiciones. De esta tesis surge que el método para detectar DIF más recomendable, en cuanto al riesgo de cometer Error de Tipo I, es en condiciones generales, la prueba normal para la diferencia de los parámetros de dificultad. Si no se dispone de los recursos que involucra el estudio del DIF con los métodos de la Teoría de Respuesta al Ítem, puede recurrirse al método ETS. Éste tiene el mismo costo computacional que el método MH, pero falla mucho menos en la detección errónea del DIF.